COVARIATE CLASSIFICATION USING DEPENDENT SAMPLES

by

Shibdas Bandyopadhyay
University of Minnesota

Technical Report No. 390
July, 1981

Abstract

In this study, the two population classification problem using dependent samples is extended when covariates are available for classification. Also, the model considered here includes the parameter structures relevant to growth models.

Key words and phrases: Two population classification, normal distribution, discriminators, covariates, dependent samples, likelihood ratio rules, plug-in rules, distribution of classification statistics.

American Mathematical Society subject classifications: 62H30, 62H10.

1. Introduction

Consider a two population classification problem where the two populations π_1 and π_2 are two distinct time points t_1 and t_2 respectively. Let ω be an experimental unit and $U_t \equiv U(\omega)$ be a $(p+q)x1$ vector of observations on the unit ω, observed at time t. The joint distribution of U_{t_1} and U_{t_2} is assumed (Bandyopadhyay, 1977, 1978) to be $2(p+q)$-variate normal, given by

$$(1) \quad \begin{pmatrix} U_{t_1} \\ U_{t_2} \end{pmatrix} \sim N_{2(p+q)}\left[\begin{pmatrix} m_1 \\ m_2 \end{pmatrix} , \begin{pmatrix} \Sigma & \rho\Sigma \\ \rho\Sigma & \Sigma \end{pmatrix}\right] , \ \rho^2 < 1$$

where m_1 and m_2 are $(p+q)x1$ vectors and Σ is a $(p+q)x(p+q)$ positive definite matrix.

In classification problems with covariates, components of U_t consists of X_t, a $px1$ vector of discriminators having unequal expectations μ_1 and μ_2 at t_1 and t_2 respectively, and a $qx1$ vector of covariates Y_t, having same expectation δ both at t_1 and t_2, i.e.

$$U_t = \begin{pmatrix} X_t \\ Y_t \end{pmatrix} , \quad m_i = \begin{pmatrix} \mu_i \\ \delta \end{pmatrix} , \ (i=1,2) .$$

Corresponding partition of Σ is written as

$$\Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} .$$

The problem is to classify ω into π_1 or π_2, i.e. to decide if $t=t_1$ or $t=t_2$. When parameters in (1) are not completely known, information about them is obtained from a random sample of N units $\omega_1, \omega_2, \ldots, \omega_N$, each observed both at t_1 and t_2 with $U_{i\alpha}$ as the U-observation on the unit ω_α observed at time t_i, $(i=1,2;\ \alpha=1,2,\ldots,N)$. Then $(U_{1\alpha}'\ U_{2\alpha}')'$, $(\alpha=1,2,\ldots,N)$ are i.i.d. having the common distribution given by (1). In this study we obtain classification rules and the distributions of the associated classification statistics.

For $q=0$, some aspects of the problem have been studied by the author (1977, 1978); also, the effect of ρ on the probability of correct classification is studied (1979) when $p=1$. When $\rho=0$ in (1), the problem reduces to the one studied by Cochran and Bliss (1948). When both $q=0$ and $\rho=0$, it becomes standard equal sample classification problem. If we condition on the covariates, as will be seen in the next section, the conditional model will extend the growth models considered by Brown (1947), and the model will be similar in structure to the growth models of Burnaby (1966) and Rao (1966).

Let $\bar{X}_i$ and $\bar{Y}_i$ denote the sample mean vectors of $X_{i\alpha}$, $(\alpha=1,2,\ldots,N)$ and $Y_{i\alpha}$, $(\alpha=1,2,\ldots,N)$ respectively. Define,

$$S_{x_i y_j} = \sum_{\alpha=1}^{N} (X_{i\alpha} - \bar{X}_i)(Y_{j\alpha} - \bar{Y}_j)'$$

and similarly $S_{x_i x_j}$ and $S_{y_i y_j}$,

$$P_{xy} = (1-\rho^2)^{-1}[S_{x_1y_1} - \rho S_{x_1y_2} - \rho S_{x_2y_1} + S_{x_2y_2}]$$

and similarly P_{xx} and P_{yy},

$$T_{xy} = S_{x_1y_1} + S_{x_1y_2} + S_{x_2y_1} + S_{x_2y_2}$$

and similarly T_{xx} and T_{yy},

$$R_{xy}=S_{x_1y_1}-S_{x_1y_2}-S_{x_2y_1}+S_{x_2y_2}$$

and similarly R_{xx} and R_{yy} .

P^*, T^*, and R^* matrices are defined the same way as P, T, and R respectively, replacing $\bar{X}_i$ by μ_i and $\bar{Y}_i$ by δ. Residual matrix $\Sigma_{x.y}$ is defined as

$$\Sigma_{x.y} = \Sigma_{xx}-\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}$$

and similarly $P_{x.y}$, $P^*_{x.y}$, $T_{x.y}$, $T^*_{x.y}$, $R_{x.y}$, and $R^*_{x.y}$. Also define,

$$||a;B||=a'B^{-1}a \; .$$

2. Classification Rules

Let H_i denote the hypothesis that $t=t_i$, and L_i denote the likelihood of the sample and U_t under H_i, $(i=1,2)$. A likelihood ratio (LR) rule is to accept H_1 if, and only if,

$$\Lambda = \sup L_1/\sup L_2$$

is large, where sup is the supremum and it is taken over the unknown parameters involved in the likelihoods.

It follows from (1) that the conditional joint distribution of X_{t_1} and X_{t_2} given Y_{t_1} and Y_{t_2} is 2p-variate normal given by

$$(2) \quad \begin{pmatrix} X_{t_1} \\ X_{t_2} \end{pmatrix} \Bigg| \begin{pmatrix} Y_{t_1} \\ Y_{t_2} \end{pmatrix} \sim N_{2p} \left[\begin{pmatrix} \mu_1+\beta(Y_{t_1}-\delta) \\ \mu_2+\beta(Y_{t_2}-\delta) \end{pmatrix}, \begin{pmatrix} \Sigma_{x.y} & \rho\Sigma_{x.y} \\ \rho\Sigma_{x.y} & \Sigma_{x.y} \end{pmatrix} \right] ,$$

where $\beta=\Sigma_{xy}\Sigma_{yy}^{-1}$. It may be noted that the conditional model (2) extends the model studied by Brown (1947) and it is similar in structure, relevant to growth models of Burnaby (1966) and Rao (1966) when $\rho=0$.

The likelihood of U_t and $U_{i\alpha}$, $(i=1,2;\alpha=1,2,\ldots,N)$ under H_1 may be written as

$$L_1 = f_Y(2\pi)^{-p(2N+1)/2}(1-\rho^2)^{-pN}|\Sigma_{x.y}|^{-(2N+1)/2}$$

$$\cdot\ \exp[-(1/2)(1-\rho^2)^{-1}Q_{21}-(1/2)Q_{11}-(1/2)Q_1]\ ,$$

and under H_2 ,

$$L_2 = f_Y(2\pi)^{-p(2N+1)/2}(1-\rho^2)^{-pN}|\Sigma_{x.y}|^{-(2N+1)/2}$$

$$\cdot\ \exp[-(1/2)(1-\rho^2)^{-1}Q_{12}-(1/2)Q_{22}-(1/2)Q_2]\ ,$$

where, for $i=1,2$ and $j=1,2$

$$Q_{ij} = \sum_{\alpha=1}^{N} \| x_{i\alpha}-\mu_i-\rho(x_{j\alpha}-\mu_j)-\boldsymbol{\beta}\{Y_{i\alpha}-\delta-\rho(Y_{j\alpha}-\delta)\};\Sigma_{x.y}\|\ ,\quad i\neq j$$

$$Q_{jj} = \sum_{\alpha=1}^{N} \| x_{j\alpha}-\mu_j-\boldsymbol{\beta}(Y_{j\alpha}-\delta);\Sigma_{x.y}\|\ ,$$

$$Q_j = \| x_t-\mu_j-\boldsymbol{\beta}(Y_t-\delta);\Sigma_{x.y}\|$$

and f_Y is the joint density function of Y_t and Y_i , $(i=1,2;\alpha=1,2,\ldots,N)$.

It may be useful to note that,

(3) $$(1-\rho^2)^{-1}Q_{21}+Q_{11}=(1-\rho^2)^{-1}Q_{12}+Q_{22}$$

for all ρ, μ_i, δ, $\boldsymbol{\beta}$, and $\Sigma_{x.y}$. We now consider cases depending on the knowledge of the parameters in (1) and obtain Λ.

CASE 1. All parameters are known

In this case Λ reduces to

$$\Lambda = \exp[Q_2 - Q_1)/2]$$

and hence, a LR rule is to accept H_1 if, and only if,

$$\| X_t - \mu_1 - \beta(Y_t - \delta); \Sigma_{x.y} \| - \| X_t - \mu_2 - \beta(Y_t - \delta); \Sigma_{x.y} \| < k_1 \tag{4}$$

for some constant k_1.

CASE 2. μ_i, δ, and Σ are known, ρ is unknown.

From the structures of L_1 and L_2, and using (3), it follows that a LR is to accept H_1 if, and only if (4) holds.

CASE 3. μ_i, δ, and ρ are known, Σ is unknown.

Under H_i, maximum likelihood estimator (m.l.e.) of β is given by

$$\hat{\beta}_i^* = [P_{xy}^* + (X_t - \mu_i)(Y_t - \delta)'][P_{yy}^* + (Y_t - \delta)(Y_t - \delta)']^{-1} ,$$

which may be rewritten as

$$\hat{\beta}_i^* = \hat{\beta}^* + E_i^*$$

where,

$$\hat{\beta}^* = P_{xy}^* \; P_{yy}^{*\,-1}$$

and, for $\theta = \| Y_t - \delta;\ P_{yy}^* \|$,

$$E_i^* = (1+\theta)^{-1}[X_t - \mu_i - \beta(Y_t - \delta)](Y_t - \delta)' P_{yy}^{*\,-1} , \quad (i=1,2).$$

Long, but straightforward simplification yields the m.l.e. of $\Sigma_{x.y}$, under H_i, as

$$(2N+1)\hat{\Sigma}^{(i)}_{x.y} = P^*_{x.y} + (1+\theta)^{-1}[X_t-\mu_i-\hat{\beta}^*(Y_t-\delta)][X_t-\mu_i-\hat{\beta}^*(Y_t-\delta)]'.$$

Thus, a LR rule accepts H_1 if, and only if,

$$\frac{1+(1+\theta)^{-1}\ \|X_t-\mu_1-\hat{\beta}^*(Y_t-\delta);P^*_{x.y}\|}{1+(1+\theta)^{-1}\ \|X_t-\mu_2-\hat{\beta}^*(Y_t-\delta);P^*_{x.y}\|} < k_3 \tag{5}$$

for some constant k_3.

CASE 4. Σ and ρ are known, μ_i and δ are unknown.

First we maximize the likelihood under H_1. Q_{21} is minimized for $\mu_2 = \hat{\mu}_2$ given by

$$\hat{\mu}_2 = \bar{X}_2-(\bar{X}_1-\mu_1)-\beta\{\bar{Y}_2-\delta-\rho(\bar{Y}_1-\delta)\}$$

and thus

$$\inf_{\mu_2} Q_{21} = \sum_{\alpha=1}^{N} \|X_{2\alpha}-\bar{X}_2-\rho(X_{1\alpha}-\bar{X}_1)-\beta\{\bar{Y}_{2\alpha}-\bar{Y}_2-\rho(Y_{1\alpha}-\bar{Y}_1\};\Sigma_{x.y}\|$$

which is free of μ_1 and δ. $Q_{11}+Q_1$ is minimized for $\mu_1 = \hat{\mu}_1$ given by

$$\hat{\mu}_1 = [N\{\bar{X}_1-\beta(\bar{Y}_1-\delta)\}+X_t-\beta(Y_t-\delta)]/(N+1)$$

and thus, after some simplification,

$$\inf_{\mu_1}(Q_{11}+Q_1) = \sum_{\alpha=1}^{N}\|X_{1\alpha}-\bar{X}_1-\beta(Y_{1\alpha}-\bar{Y}_1);\ \Sigma_{x.y}\|$$
$$+ [N/(N+1)]\,\|X_t-\bar{X}_1-\beta(Y_t-\bar{Y}_1);\ \Sigma_{x.y}\| \ ,$$

which does not involve δ. Similarly the likelihood is maximized, under H_2, when sufficies 1 and 2 are interchanged. Again using (3), a LR rule is to accept H_1 if, and only if,

$$(6) \qquad \| X_t-\bar{X}_1-\beta(Y_t-\bar{Y}_1);\ \Sigma_{x.y}\| - \| X_t-\bar{X}_2-\beta(Y_t-\bar{Y}_2);\ \Sigma_{x.y}\| < k_4$$

for some constant k_4.

CASE 5. Σ is known, μ_i, δ, and ρ are unknown.

From the structures of L_1 and L_2, and using (3), it follows that a LR rule is to accept H_1 if, and only if (6) holds.

CASE 6. ρ is known, μ_i, δ, and Σ are unknown.

Likelihood is maximized for μ_1 and μ_2 given under case 4, when β is replaced, under H_i, by $\hat{\beta}_i$ given by

$$\hat{\beta}_i = \hat{\beta} + E_i$$

where,

$$\hat{\beta} = P_{xy}\, P_{xy}^{-1}$$

and, for $\theta_i = \| Y_t-\bar{Y}_i; P_{yy}\|$,

$$E_i = [\theta_i+(N+1)/N]^{-1}[X_t-\bar{X}_i-\hat{\beta}(Y_t-\bar{Y}_i)](Y_t-\bar{Y}_i)' P_{yy}^{-1}, \quad (i=1,2) .$$

The m.l.e. of $\Sigma_{x.y}$, under H_i, now is given by (i=1,2)

$$(2N+1)\ \hat{\Sigma}^{(i)}_{x.y} = P_{x.y}$$

$$+ [\theta_i+(N+1)/N]^{-1} [X_t-\bar{X}_i-\hat{\beta}(Y_t-Y_i)]\ (Y_t-\bar{Y}_i)'\ P^{-1}_{yy}\ .$$

Thus, once again using (3), a LR rule is to accept H_1 if, and only if,

$$\frac{1+[\theta_1+(N+1)/N]^{-1}\ \| X_t-\bar{X}_1-\hat{\beta}(Y_t-\bar{Y}_1);P_{x.y}\|}{1+[\theta_2+(N+1)/N]^{-1}\ \| X_t-\bar{X}_2-\hat{\beta}(Y_t-\bar{Y}_2);P_{x.y}\|} < k_6 \tag{7}$$

for some constant k_6.

When both ρ and Σ are unknown, $\sup L_i$ cannot be written in a closed form as one needs to solve for ρ and Σ by iterative methods. In such cases plug-in likelihood ratio (PLR) rules (Wald, 1944) are obtained by replacing the unknown parameters involved in the LR rule (4) with respective consistent and unbiased estimators, based only on the sample.

CASE 7. μ_i and δ are known, ρ and Σ are unknown.

In this case, an unbiased consistent estimator of β is

$$\tilde{\beta} = (1/2)(\tilde{\beta}^*_1+\tilde{\beta}^*_2)$$

where, $\tilde{\beta}^*_1 = T^*_{xy}T^{*\ -1}_{yy}$ and $\tilde{\beta}^*_2 = R^*_{xy}R^{*\ -1}_{yy}$. An unbiased consistent estimator of $\Sigma_{x.y}$ is

$$\tilde{\Sigma}^*_{x.y} = (1/4)(\tilde{\Sigma}^*_1+\tilde{\Sigma}^*_2)$$

where, $\tilde{\Sigma}^*_1 = (N-q)^{-1}\ T^*_{x.y}$ and $\tilde{\Sigma}^*_2 = (N-q)^{-1}\ R^*_{x.y}$. So a PLR rule accepts H_1 if, and only if,

(8) $$\| X_t-\mu_1-\tilde{\beta}^*(Y_t-\delta);\tilde{\Sigma}^*_{x.y}\|-\|X_t-\mu_2-\tilde{\beta}^*(Y_t-\delta);\tilde{\Sigma}^*_{x.y}\| < k_7$$

for some constant k_7.

CASE 8. All parameters are unknown.

In this case a set of unbiased consistent estimators of μ_i, δ, β, and $\Sigma_{x.y}$ are

$$\tilde{\mu}_i = \bar{X}_i \ , \ (i=1,2) \ ,$$

$$\tilde{\delta} = (1/2)(\bar{Y}_1+\bar{Y}_2) \ ,$$

$$\tilde{\beta} = (1/2)(\tilde{\beta}_1+\tilde{\beta}_2) \ ,$$

and

$$\tilde{\Sigma}_{x.y} = (1/4)\ [\tilde{\Sigma}_1+\tilde{\Sigma}_2]$$

where, $\tilde{\beta}_1 = T_{xy}T_{yy}^{-1}$, $\tilde{\beta}_2 = R_{xy}R_{yy}^{-1}$, $\tilde{\Sigma}_1 = (N-q-1)^{-1}\ T_{x.y}$, and $\tilde{\Sigma}_2 = (N-q-1)^{-1}\ R_{x.y}$. Hence, a PLR rule accepts H_1 if, and only if,

(9) $$\| X_t-\bar{X}_1-\tilde{\beta}(Y_t-\bar{Y}_1);\tilde{\Sigma}_{x.y}\| - \| X_t-\bar{X}_2-\tilde{\beta}(Y_t-Y_2);\tilde{\Sigma}_{x.y}\| < k_8$$

for some constant k_8.

3. Distributions

In this section the distributions of some classification statistics are considered. Though no attempt is made to write the explicit forms of the densities, these statistics are suitably reduced to forms where standard results are applicable.

In what follows, the <u>conditional</u> distributions, <u>given</u> the covariates $Y_{i\alpha}$, $(i=1,2;\alpha=1,2,\ldots,N)$ and Y_t, are obtained when H_1 is true.

CASES 1 AND 2. In these cases, the classification statistic associated with the LR rule given by (4) is

$$T_1 = [2X_t-\mu_1-\mu_2 -2\beta(Y_t-\delta)]' \Sigma_{x.y}^{-1}(\mu_1-\mu_2) .$$

It follows from (2) that T_1 is univariate normal with mean $(-\Delta^2)$ and variance $4\Delta^2$, where

$$\Delta^2 = \| \mu_1-\mu_2;\Sigma_{x.y}\| . \tag{10}$$

CASE 3. Consider the statistic

$$T_3 = (1+\theta)^{-1}[2X_t-\mu_1-\mu_2-2\hat{\beta}^*(Y_t-\delta)]'P_{x.y}^{*-1}(\mu_1-\mu_2) .$$

This statistic is associated with the LR rule given by (5) when $k_3 = 1$. To obtain the (conditional) distribution of T_3, we first note that $(1-\rho^2)^{-1/2}[X_{1\alpha}-\mu_1-\rho(X_{2\alpha}-\mu_2)]$ and $(X_{2\alpha}-\mu_2)$, $(\alpha=1,2,\ldots,N)$ are 2N mutually (conditionally) independent normal variates, each having the same covariance matrix $\Sigma_{x.y}$. The mean of $(X_{1\alpha}-\mu_1)-\rho(x_{2\alpha}-\mu_2)$ is $\beta[Y_{1\alpha}-\delta-\rho(Y_{2\alpha}-\delta)]$ and the mean of $(X_{2\alpha}-\mu_2)$ is $\beta(Y_{2\alpha}-\delta)$. Thus [Anderson (1958), theorem 8.2.2] $\hat{\beta}^* = P_{xy}^*P_{yy}^{*-1}$ and $P_{x.y}^*$ are (conditionally) independent. Also,

$$\hat{\beta}^*(Y_t-\delta) \sim N_p[\beta(Y_t-\delta),\ \theta\Sigma_{x.y}] ,$$

and

$$P^*_{x.y} \sim W_p[\Sigma_{x.y},\ 2N-q]\ ,$$

where $W_p[\tau,k]$ denotes p-variate Wishart distribution with k degrees of freedom and matrix parameter τ. Thus, $[2X_t-\mu_1-\mu_2-2\hat{\beta}^*(Y_t-\delta)]$ is p-variate normal with mean $(\mu_1-\mu_2)$ and covariance matrix $4(1+\theta)\Sigma_{x.y}$, and (conditionally) independent of $P^*_{x.y}$.

It may be observed that the (conditional) distribution of T_3 depends only on the parameter Δ^2 given by (10). We shall denote the (conditional) density of T_3 as $f_\Delta(\cdot|p,2N-q)$.

CASES 4 AND 5. The classification statistic associated with the LR rule given by (6) may be written as

$$T_4 = (c_1c_2)^{1/2}\ V_1'\Sigma^{-1}_{x.y}V_2$$

where,

$$V_1 = c_1^{-1/2}[2(X_t-\beta Y_t)-(\bar{X}_1-\beta\bar{Y}_1)-(\bar{X}_2-\beta\bar{Y}_2)]\ ,$$

$$V_2 = c_2^{-1/2}[(\bar{X}_2-\beta\bar{Y}_2)-(\bar{X}_1-\beta\bar{Y}_1)]\ ,$$

$$c_1 = 4+2(1+\rho)/N, \quad \text{and} \quad c_2 = 2(1-\rho)/N\ .$$

It follows from (2) that V_1 and V_2 are (conditionally) independent p-variate normal, each having the covariance matrix $\Sigma_{x.y}$. We may write,

$$2(c_1c_2)^{-1/2}T_4 = (1/2)[\|V_1+V_2;\Sigma_{x.y}\| - \|V_1-V_2;\Sigma_{x.y}\|]\ .$$

Now, $(1/2)\|V_1+V_2;\Sigma_{x.y}\|$ has a non-central chi-square distribution with p-degrees of freedom and non-centrality parameter $(1/2)\Delta^2[c_1^{-1/2}+c_2^{-1/2}]$, and $(1/2)\|V_1-V_2;\Sigma_{x.y}\|$ also has a non-central chi-square distribution with p-degrees of freedom and non-centrality parameter $(1/2)\Delta^2[c_1^{-1/2}-c_2^{-1/2}]$. Since V_1+V_2 and V_1-V_2 are (conditionally) independent, the two chi-squares are also independent. The density of the difference of two independent non-central chi-squares is given by John (1960).

CASE 6. Consider the statistic

$$T_6 = V_1'P_{x.y}^{-1}V_1-V_2'P_{x.y}^{-1}V_2 ,$$

where

$$V_i = [\theta_i+(N+1)/N]^{-1/2} [(X_t-\bar{X}_i)-\hat{\beta}(y_t-\bar{Y}_i)] , (i-1,2) .$$

This statistic is associated with the LR rule given by (7) when $k_6=1$. Analysis similar to case 3 above yields, in particular, that $\bar{X}_1-\rho\bar{X}_2$, $\bar{X}_2$, $\hat{\beta}$, and $P_{x.y}$ are (conditionally) mutually independent. For ξ given by

$$\xi = [\theta_1+(N+1)/N]^{-1/2} [\theta_2+(N+1)/N]^{-1/2} [(N+\rho)/N+(Y_t-\bar{Y}_1)'P_{yy}^{-1}(Y_t-\bar{Y}_2)] ,$$

$$(2+\xi)^{-1/2}(V_1+V_2) \sim N_p[(2+\xi)^{-1/2}(\mu_1-\mu_2) , \Sigma_{x.y}] ,$$

$$(2-\xi)^{-1/2}(V_1-V_2) \sim N_p[-(2-\xi)^{-1/2}(\mu_1-\mu_2) , \Sigma_{x.y}] ,$$

$$P_{x.y} \sim W_p[\Sigma_{x.y} , 2N-2-q] ,$$

and these three (conditional) distributions are mutually independent. Hence the (conditional) distribution of $(4-\xi^2)^{-1/2}T_6 = (4-\xi^2)^{-1/2}(V_1+V_2)'P^{-1}_{x.y}(V_1-V_2)$ may be obtained from Sitgreaves (1962). The density, which depends on the parameter Δ^2, will be denoted by $g_\Delta(\cdot|p,2N-2-q)$.

CASE 7. The classification statistic associated with the PLR rule given by (8) is

$$T_7 = [2X_t-\mu_1-\mu_2-2\tilde{\beta}^*(Y_t-\delta)]'\tilde{\Sigma}^{*-1}_{x.y}(\mu_2-\mu_1) .$$

It follows from (2) that $X_{1\alpha}+X_{2\alpha}$ and $X_{1\alpha}-X_{2\alpha}$, $(\alpha=1,2,\ldots, N)$ are (conditionally) mutually independent p-variate normal given by

$$X_{1\alpha}+X_{2\alpha} \sim N_p[\mu_1+\mu_2+\beta\{(Y_{1\alpha}-\delta)+(Y_{2\alpha}-\delta)\},\ 2(1+\rho)\Sigma_{x.y}] ,$$

$$X_{1\alpha}+X_{2\alpha} \sim N_p[\mu_1+\mu_2+\beta\{(Y_{1\alpha}-\delta)-(Y_{2\alpha}-\delta)\},\ 2(1-\rho)\Sigma_{x.y}] .$$

Applying theorem 8.3.3 of Anderson (1958) separately to $X_{1\alpha}+X_{2\alpha}$, $(\alpha=1,2,\ldots, N)$, and to $X_{1\alpha}-X_{2\alpha}, (\alpha=1,2,\ldots N)$, it is seen that $[2X_t-\mu_1-\mu_2-2\tilde{\beta}^*(Y_t-\delta)]$ is (conditionally) p-variate normal with mean $(\mu_1-\mu_2)$ and covariance matrix $[4+2(1+\rho)\| Y_t-\delta;T^*_{yy}\| +2(1-\rho)\| Y_t-\delta;R^*_{yy}\|]\Sigma_{x.y}$, and (conditionally) independent of $\tilde{\Sigma}^*_{x.y}$. Also, $[2(1+\rho)]^{-1} T^*_{x.y}$ and $[2(1-\rho)]^{-1} R^*_{x.y}$ are (conditionaly) independent $W_p[\Sigma_{x.y}, N-q]$. Thus, it follows from Bandyopadhyay (1978) that the density function of $[(N-q)/\lambda]\tilde{\Sigma}^*_{x.y}$ is $\sum_{j=0}^{\infty} p_j(N-q,\lambda)h(\cdot|\Sigma_{x.y},2N+2j-2q)$, where $\lambda=(1-|\rho|)/2$ and $p_j(N,\lambda) = [(1-2\lambda)\lambda^{N/2}\Gamma(N/2+j)][\Gamma(N/2)(1-\lambda)^{N/2+j}j!]^{-1}$ with $\sum_{j=0}^{\infty} p_j(N,\lambda) = 1$, $h(\cdot|\tau,k)$ being the density function of $W_p[\tau,k]$. Thus the form of the density of T_7 is $\sum_{j=0}^{\infty} p_j(N-q,\lambda)\ f_\Delta(\cdot|p,2N+2j-2q)$, where f_Δ is defined under case 3.

CASE 8. The classification statistic associated with the PLR rule given by (9) is

$$T_8 = \|(X_t-\bar{X}_1)-\tilde{\beta}(Y_t-\bar{Y}_1);\ \tilde{\Sigma}_{x.y}\| - \|(X_t-\bar{X}_2)-\tilde{\beta}(Y_t-\bar{Y}_2);\ \tilde{\Sigma}_{x.y}\| .$$

Combining analysis similar to case 6 with set-up of case 7, separately applying to $X_{1\alpha} + X_{2\alpha}$, $(\alpha = 1, 2, \ldots, N)$, and to $X_{1\alpha} - X_{2\alpha}$, $(\alpha = 1, 2, \ldots, N)$, the form of the density of T_8 is $\sum_{j=0}^{\infty} p_j(N-1-q,\lambda)g_\Delta(\cdot|p,\ 2N + 2j - 2q - 2)$, where g_Δ is define under case 6.

REFERENCES

[1] Anderson, T. W. (1958), An Introduction to Multivariate Statistical Analysis, Wiley, New York.

[2] Bandyopadhyay, S. (1977), Probability inequalities involving estimates of probability of correct classification using dependent samples, Sankhyā (B), 39, 145-150.

[3] Bandyopadhyay, S. (1978), Classification, distribution and admissibility in autoregressive process, Austral. J. Statist., 18, 514-528.

[4] Bandyopadhyay, S. (1979), Two population classification in Gaussuan process, J. Statist. Planning and Inf., 3, 225-233.

[5] Brown, G. W. (1974), Discriminant functions, Ann. Math. Statist., 18, 514-528.

[6] Burnaby, T. P. (1966), Growth invariant discriminant function and generalized distances, Biometrics, 22, 96-110.

[7] Cochran, W. G. and Bliss, C. I. (1948), Discriminant functions with covariance, Ann. Math. Statist., 19, 151-176.

[8] John, S. (1960), On some classification statistics, Sankhyā, 22, 310-316 (correction: 23, 308).

[9] Rao, C. R. (1966), Discriminant function between composite hypothesis and related problems, Biometrika, 53, 339-345.

[10] Sitgreaves, R. (1961), Some results on distributions of W-classification statistics, Studies in Item Analysis and Prediction (Ed. H. Soloman), Stanford University Press, Stanford.

[11] Wald, A. (1944), On a statistical problem arising in the classification of an individual into one of two groups, Ann. Math. Statist., 15, 145-162.